

Data Production in ECS

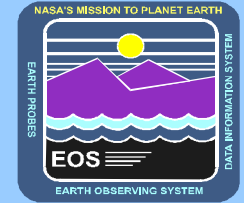
Planning and Data Processing Subsystems

William Knauss

4-5 June 1997



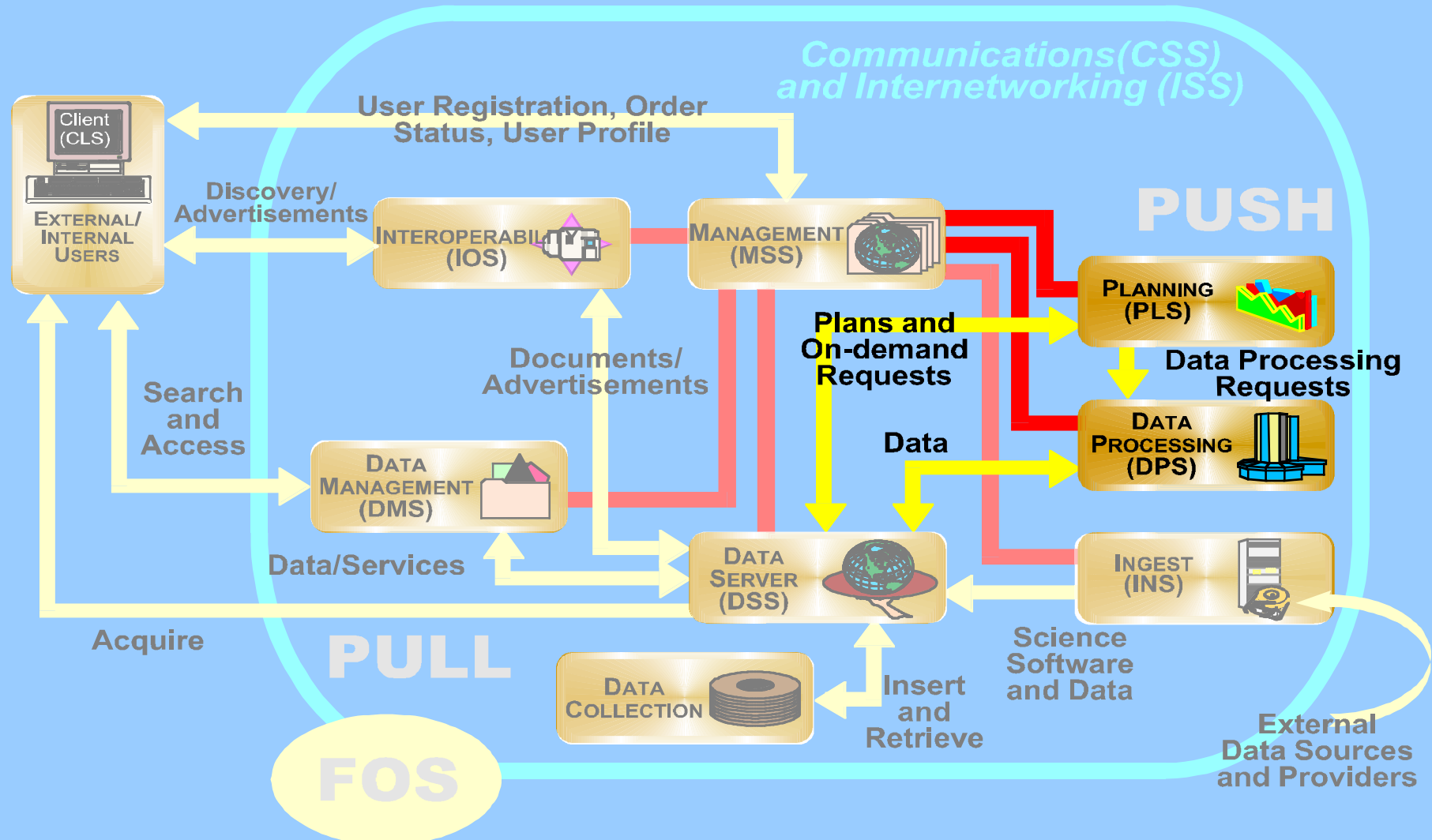
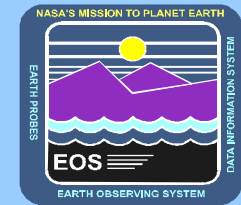
Agenda



- ◆ **The Big Picture**
- ◆ **PGEs and Production Rules**
- ◆ **Science Software Integration & Test (SSI&T)**
- ◆ **Planning and Data Processing Glossary**
- ◆ **Production Planning**
- ◆ **Data Processing**
- ◆ **Special Topics**
 - **Production Rules**
 - **On-Demand Processing**
 - **Cross-DAAC Planning**

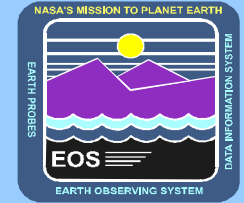


PDPS Context





The Big Picture

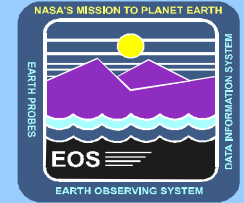


ECS is a data production and distribution system

- ✦ **Scientists write software which processes satellite data to produce products**
- ✦ **ECS provides the environment in which that software runs through the ECS SDP Toolkit, PLS and DPS**



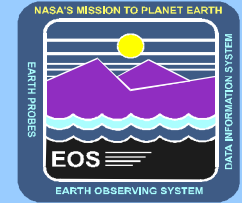
The Scientist's Software



- ◆ **PGE (Product Generation Executive) - Science software consisting of a script and one or more binary executables.**
- ◆ **SDP Toolkit - ECS Library of routines used by PGEs. Really the Interface between PGEs and ECS. Provides I/O routines, does geographic coordinate conversions, etc.**



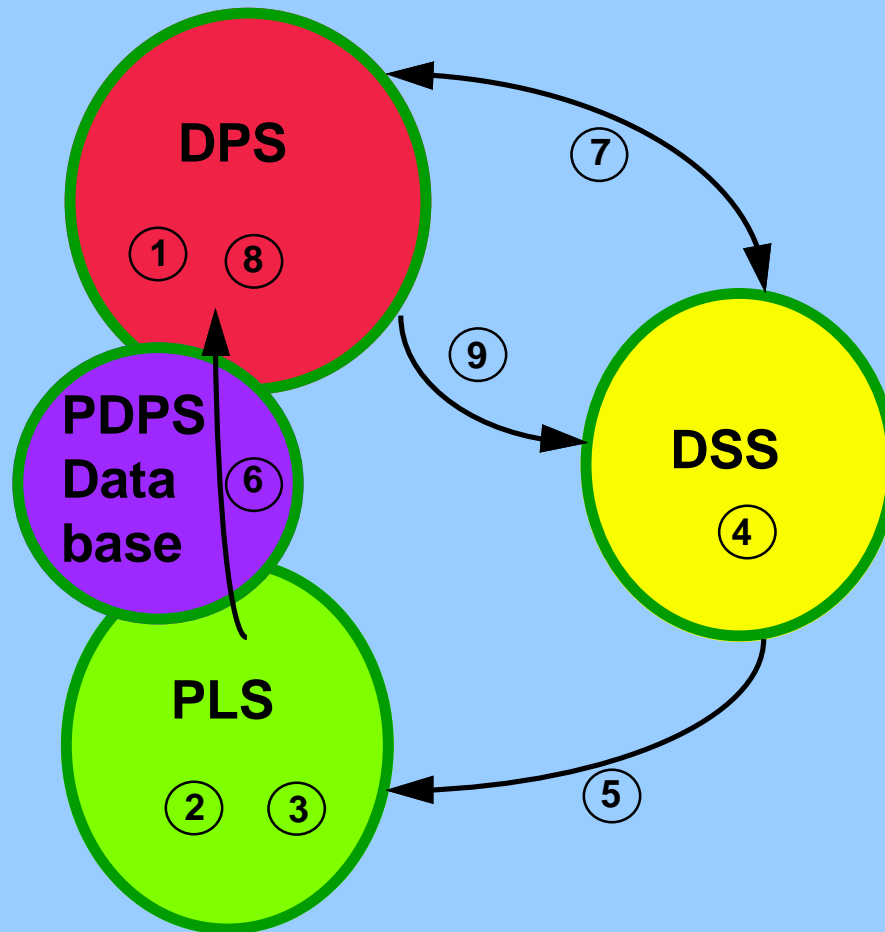
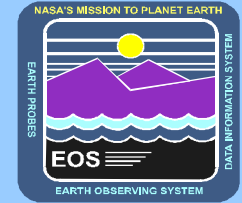
ECS Environment to Support PGE Execution



- ✦ **PLS (Planning Subsystem) - Subsystem responsible for planning PGE runs. Major COTS: Sybase, Hughes Delphi Class Libraries.**
- ✦ **DPS (Data Processing Subsystem) - Subsystem responsible for executing PGEs. Major COTS: Sybase, AutoSys.**
- ✦ **DSS (Data Server Subsystem) - Subsystem responsible for archiving and providing access to the science data.**



Simplified Data Production



1. Capture PGE info. in PDPS Database

2. Enter Production Requests in PLS

3. Create/activate Production Plan in PLS

4. Data required becomes available at DSS

5. DSS sends Subscription notice to PLS

6. PLS releases jobs to DPS

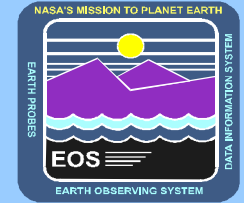
7. DPS acquires inputs from DSS

8. DPS runs jobs

9. DPS inserts outputs to DSS.



Levels of Data

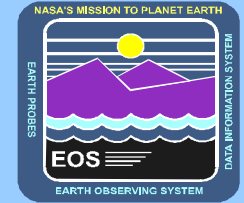


Level 0 - Raw instrument data at original resolution, time ordered, with duplicate packets removed.

Level 1 - Level 0 data located to a coordinate system, and packaged with needed ancillary and engineering data (Level 1A) and radiometrically corrected and calibrated data in physical units at full instrument resolution (Level 1B).



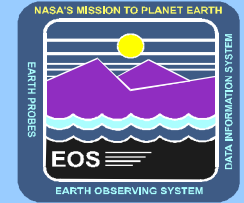
Levels of Data (cont.)



Level 2 - Retrieved environmental variables (e.g., ocean wave height, soil moisture, ice concentration) at the same location and similar resolution as the Level 1 source data.

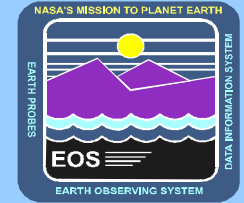
Level 3 - Data or retrieved environmental variables that have been spatially and/or temporally resampled (i.e., derived from Level 1 or Level 2 data products). Such resampling may include averaging and compositing.

Level 4 - Model output and/or variables derived from lower level data which are not directly measured by the instruments. For example, new variables based upon a time series of Level 2 or Level 3 data.



Types of Data Processing

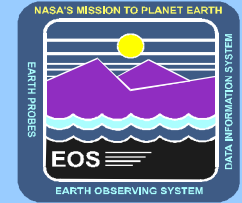
- ✦ **Routine Processing** - Pre-defined software production processing which is periodic and keyed to data arrival.
- ✦ **On-Demand Processing** - Ad-hoc processing which is (usually) initiated by a user request.
- ✦ **Reprocessing** - Very similar to Routine Processing. As algorithms improve, large scale reprocessing programs will be undertaken to improve data quality.



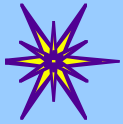
Science Software Integration & Test (SSI&T)



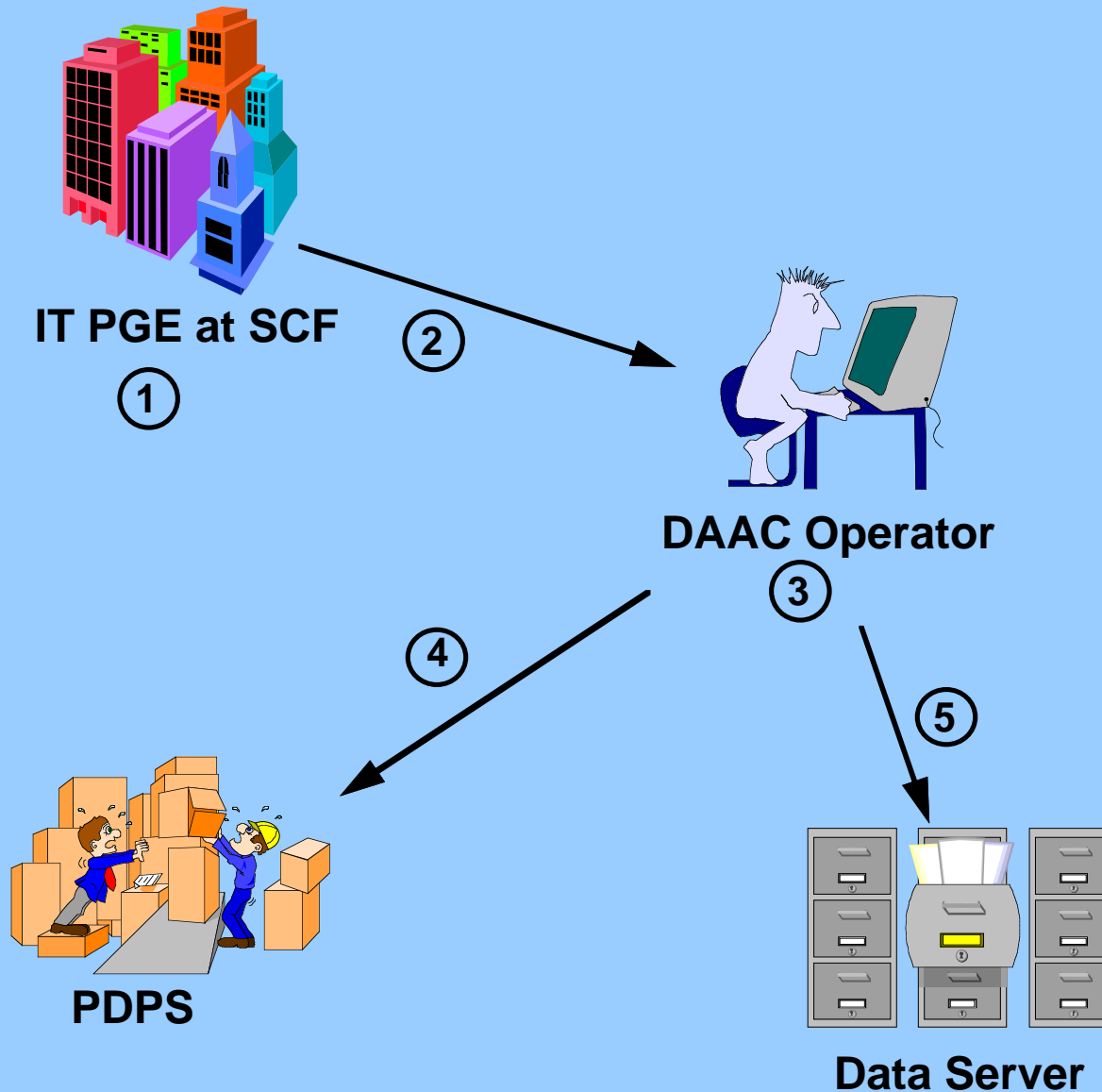
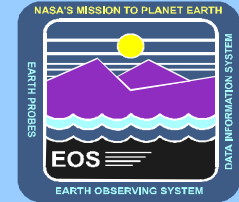
Science Software Integration & Test (SSI&T)



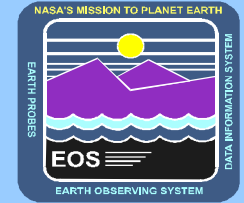
- ✦ **SSI&T is the process PGEs go through to bring them into ECS and make them available to run**
- ✦ **SSI&T Consists of several steps including:**
 - **Prohibited function check**
 - **Process Control File (PCF) template verification**
 - **Build executables**
 - **Memory Leak detection**
 - **PDPS Database population**
- ✦ **Instrument Teams provide PGEs to SSI&T Personnel at DAACs**
- ✦ **SSI&T is done by DAAC science personnel using SDE procedure**



SSI&T Process for Adding PGEs



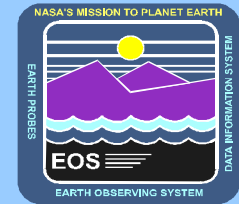
- ① IT Creates PGE
- ② Software (and related info.) transmitted to DAAC
- ③ DAAC operator Performs SSI&T
- ④ PGE info. is entered in PDPS Database
- ⑤ Science Software Archive Package sent to Data Server



Planning and Data Processing Glossary



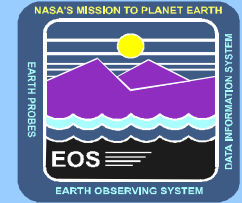
Planning and Data Processing Glossary



- ✦ **PR (Production Request)** - The mechanism for a Production Planner to request products to be generated. A PR normally specifies a PGE and the data collection time range for data to be produced. The PGE inherently specifies the product to be produced. Will lead to the creation of multiple DPRs.



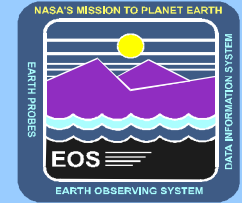
Planning and Data Processing Glossary (continued)



- ◆ **DPR (Data Processing Request) - Generated by Planning Subsystem using a PR and information from PGE profile in PDPS database. One DPR corresponds to a single Product Generation Executive (PGE) to be executed. A DPR includes a PGE, input data granule(s), output data granule and archive location, planned start/end execution times, and priority. DO NOT CONFUSE WITH ASTER DPRs!**



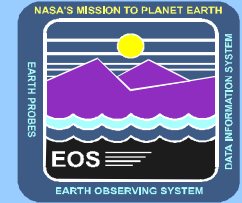
Planning and Data Processing Glossary (continued)



- ◆ **Production Plan** - A predicted order of DPRs which embodies the processing objectives for a particular time period.
 - **CANDIDATE PLAN** - The result of “what if” planning by the Production Planner.
 - **ACTIVE PLAN** - Is a Candidate Plan that has been selected and placed into execution by the Production Planner.



Planning and Data Processing Glossary (continued)



- ✦ **PCF (Process Control File) - File created by DPS from PGE template and used by the SDP toolkit. Provides PGE with runtime parameters and links logical file numbers used by PGE with physical filenames of data files.**